Monocular Visual Localization via Conic Geometric Cues for Astrobee Free-Flyers

Minji Kim[1], Suyoung Kang[2], Seonwook Yeom[1], Ryan Soussan[3], Brian Coltin[3], and Pyojin Kim[1]

***Abstract*— We present a 6-DoF monocular visual localization method for free-flying robots inside the ISS, leveraging time-invariant conic geometric cues and Manhattan world (MW) regularities in structured scenes for globally aligned poses. Existing visual localization methods for free-flying robots rely on handcrafted or learned keypoints matched against pre-built maps of the ISS, making them vulnerable to illumination changes, occlusions, and frequent environmental reconfigurations. To address these limitations, we exploit time-invariant conic primitives from circular landmarks and MW regularities in human-made environments. A Perspective-Two-Circle (P2C) solver recovers an initial pose from two detected ellipses using only their known radii. The resulting conic normals are aligned with MW lines for drift-free orientation. All poses are jointly optimized through conic bundle adjustment (CBA) with algebraic conic residuals. Experiments on Astrobee datasets show that our method outperforms state-of-the-art Astrobee localizers. It further achieves accuracy comparable to feature-map-based approaches under severe illumination changes.**

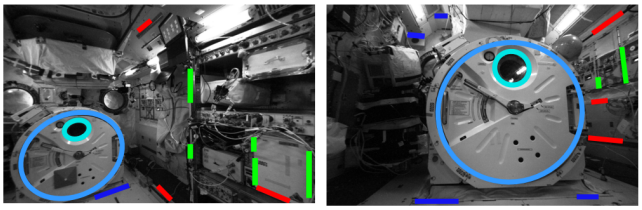
I. Introduction

Visual Localization (VL) is essential for accurate state estimation of free-flying robots such as Astrobee [1], [2], [3] and Int-Ball2 [4] operating on the ISS. However, lighting changes and frequent facility reconfiguration [5] cause inconsistencies between pre-built maps and the current scene, degrading VL performance [6] despite unchanged structural geometry.

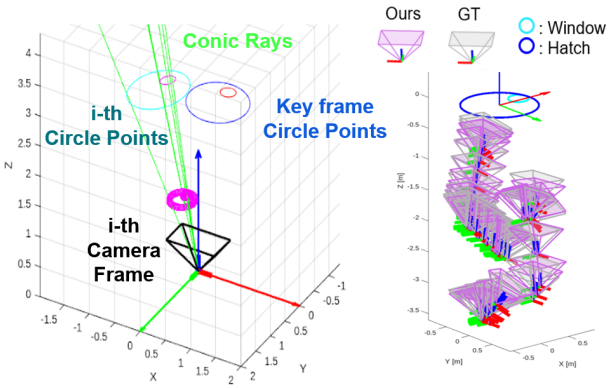
The state-of-the-art free-flyer localizer AstroLoc2 [7] improves noise robustness but still depends on pre-built feature maps [8] requiring remapping, while visual SLAM methods for Int-Ball2 and CIMON [9] based on ORB-SLAM2 [10] suffer from drift and lack global localization capability. Recent map-free localization approaches [11], [12], [13], [14] reduce map dependency but rely on large-scale training and recover only relative poses without global localization.

To address these issues, we propose a monocular visual localization method that exploits conic geometry from two coplanar circular landmarks with known physical radii. Combined with Manhattan World (MW) regularities, our method enables globally consistent 6-DoF pose estimation via a P2C solver for initialization and conic bundle adjustment (CBA) for refinement, without pre-built maps, depth sensing, or appearance-based correspondences. Our contributions are:

- We leverage conic geometry for Visual Localization (VL) via a closed-form P2C initialization and a CBA that globally optimizes algebraic ellipse residuals.
- We validate our method on ISS free-flyer datasets, outperforming state-of-the-art Astrobee localizers.
- Our method demonstrates stronger robustness to illumination changes compared to map-based methods.



(a) Input image with lines and circles



(b) Circle in the i-th camera frame (c) Localization results

Fig. 1. (a) Detected ellipses and lines on input images. (b) Circle estimation in the i-th camera frame via conic ray with key-frame circle points (c) Estimated camera trajectory (Ours, magenta) against ground truth (gray).

II. Proposed Method

We propose a map-free monocular visual localization using circular landmarks and Manhattan World regularities for 6-DoF pose estimation. Our approach operates under three assumptions: (1) the physical radii of the circular landmarks are known a priori, (2) at least two circular landmarks are simultaneously visible in the image, and (3) the normal vectors of the two circle planes are parallel, i.e., the circles lie on parallel planes The pipeline is shown in Fig. 2.

A. Perspective-Two-Circle (P2C) Solver

The P2C solver recovers the 3D center $\mathbf{o}_i$ and surface normal $\mathbf{n}_i$ of the ISS hatch and window from a single

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.RS-2024-00358374) and the GIST Future-Leading Specialized Research Project grant funded by the GIST in 2026.

[1]School of Mechanical and Robotics Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, South Korea. {minji0110,seonwookyeom,pjinkim}@gm.gist.ac.kr

[2]Manning College of Info. and Computer Sciences, University of Massachusetts Amherst, MA 01002, USA. suyoungkang@umass.edu

[3]Brian Coltin and Ryan Soussan are with the Intelligent Robotics Group (KBR Inc.), NASA Ames Research Center, Moffett Field, CA 94035, USA. {brian.coltin, ryan.soussan}@nasa.gov

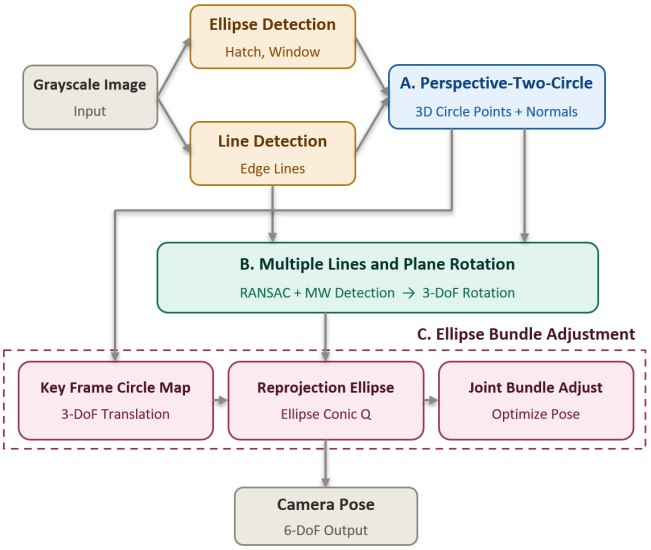


Fig. 2. Overview of our proposed three-stage pipeline: (i) ellipse and line extraction (ii) P2C solver for 3D circle geometry and pose initialization, followed by MW-based optimization (iii) Conic Bundle Adjustment (CBA) refining 6-DoF poses by minimizing algebraic conic residuals.

monocular image using known physical radii, via a YOLO-based detector followed by an ellipse fitting model [15].

The ellipse constraint yields the image-plane $\mathbf{Q}_{\text{cone}} = \mathbf{K}^\top \mathbf{Q}_{\text{ellipse}} \mathbf{K}$, forming conic ray (Fig. 1(b)), such that

$$\mathbf{X}\mathbf{K}^\top \mathbf{Q}_{\text{ellipse}} \mathbf{K}\mathbf{X}^\top = 0. \tag{1}$$

We perform eigen-decomposition on the symmetric matrix $\mathbf{Q}_{\text{cone}}$ to obtain its principal axes, yielding $U^\top \mathbf{Q}_{\text{cone}} U = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$, where $U = [\mathbf{u}_1 \ \mathbf{u}_2 \ \mathbf{u}_3]$ is an orthogonal eigenbasis and $\lambda_1, \lambda_2, \lambda_3$ are the associated eigenvalues. Given the known real-world radius R of the circular structure, the 3D circle center $\mathbf{o}_i'$ and surface normal $\mathbf{n}_i'$ are recovered in closed form from the eigenvalues and eigenvectors:

$$\mathbf{o}_i' = \begin{bmatrix} \pm R\sqrt{\dfrac{\lambda_3(|\lambda_1| - |\lambda_2|)}{|\lambda_1|(|\lambda_1| - |\lambda_3|)}} \\ 0 \\ R\sqrt{\dfrac{|\lambda_1|(|\lambda_2| + |\lambda_3|)}{\lambda_3(|\lambda_1| + |\lambda_3|)}} \end{bmatrix}, \mathbf{n}_i' = \begin{bmatrix} \pm\sqrt{\dfrac{|\lambda_1| - |\lambda_2|}{|\lambda_1| + |\lambda_3|}} \\ 0 \\ -\sqrt{\dfrac{|\lambda_2| + |\lambda_3|}{|\lambda_1| + |\lambda_3|}} \end{bmatrix}, \tag{2}$$

in the camera frame via $\mathbf{o}_i = U\mathbf{o}_i'$ and $\mathbf{n}_i = U\mathbf{n}_i'$.

Since both circles lie on the same plane, we disambiguate the symmetry by selecting the strongest normal alignment:

$$(k^*, \ell^*) = \arg\max_{k,\ell} \left(\mathbf{n}_h^{(k)}\right)^\top \mathbf{n}_w^{(\ell)}, \tag{3}$$

where $k, \ell \in \{1, 2\}$ index the candidate solutions. The $\mathbf{o}_i$ and $\mathbf{n}_i$ initialize the translation t_k, where keyframes are selected based on the minimum depth offset along the camera z-axis.

B. Multiple Lines Rotation Optimization

We exploit the MW assumption for the ISS, where P2C-derived surface normals directly yield the dominant plane normal (VD1), with VD2 estimated via RANSAC [16], and VD3 as the cross product of VD1 and VD2.

We refine the MW frame by minimizing the orthogonal distance between line inliers and their vanishing directions:

$$\theta^* = \arg\min_{\theta} \sum_{k=2}^{3} \sum_{i=1}^{M_k} \left(d_{i,k}(\theta)\right)^2, \tag{4}$$

where $d_{i,k}(\theta)$ is the orthogonal distance between the i-th line segment and the k-th vanishing point, solved via Levenberg–Marquardt, yielding the final rotation $R_k = R(\theta^*)R_k^{\text{MW}}$, where $R_k^{\text{MW}} = \begin{bmatrix}\text{VD1} & \text{VD2} & \text{VD3}\end{bmatrix}$.

C. Conic Bundle Adjustment (CBA)

After obtaining initial poses t_k and R_k, we jointly refine all camera poses through a circle-based bundle adjustment formulated from global ellipse geometry, allowing the subsequent nonlinear optimization to converge reliably. Given two sets of 3D points sampled from circles, denoted by $\mathbf{X}_{C_1}$ and $\mathbf{X}_{C_2}$, their 2D pixel reprojections under an initial camera pose $(R_k, \mathbf{t}_k)$ are given by

$$\tilde{\mathbf{x}}_i = \pi\big(R_k \, \mathbf{X}_{C_i} + \mathbf{t}_k\big), \quad i \in \{1, 2\}, \tag{5}$$

where $\pi(\cdot)$ is the perspective projection operator.

We leverage the implicit conic representation to measure the elliptic reprojection error by checking whether a reprojected point $\tilde{\mathbf{x}}_i$ lies on the detected ellipse contour. Specifically, the algebraic conic residual is defined as

$$r_{k,i}(\xi) = \tilde{\mathbf{x}}_i^\top \mathbf{Q}_i \, \tilde{\mathbf{x}}_i, \tag{6}$$

where $\xi = \{(R_k, t_k)\}_{k=1}^N$ stacks all pose parameters and $\mathbf{Q}_i$ denotes the conic matrix of the detected ellipse, enforcing geometric consistency between the reprojected 3D circle and observed 2D ellipse without explicit point correspondences. For each sampled 3D circle point $\mathbf{X}_{C_i}$ and frame k, we define the conic residuals for the two ellipses as $r_k^i(\xi)$, $i \in \{1, 2\}$, where ξ stacks the pose parameters $\{R_k, t_k\}_{k=1}^N$.

The ellipse bundle-adjustment problem is then formulated as a nonlinear least-squares optimization:

$$\xi^* = \arg\min_{\xi} \sum_{k=1}^{N} \sum_{i} \left(r_k^i(\xi)\right)^2. \tag{7}$$

We solve this using Levenberg–Marquardt with analytically derived Jacobians, yielding the optimized parameter vector ξ^* that provides refined camera poses for all frames.

III. Evaluation

We evaluate the proposed method on the public Astrobee dataset [5], covering forward, lateral, and rotation sequences in the ISS. All baseline methods are evaluated on the same subset of ellipse-visible frames, as our method requires ellipse detection as a prerequisite. We compare against Astrobee-specific methods AstroLoc [17] and AstroLoc2 [7], both of which rely on pre-built maps combined with a VIO pipeline fusing visual features, IMU, and ToF depth.

We compare against ORB-SLAM3 [18] (feature-based SLAM), DROID-SLAM [19] (learned dense correspondence), VGGT-SLAM [20] (hybrid feed-forward SLAM, pre-trained without ISS fine-tuning), COLMAP [21] (SfM-based

TABLE I

EVALUATION RESULTS OF ABSOLUTE TRAJECTORY ERROR [M] ON ASTROBEE DATASETS (FF_RETURN_JOURNEY_)

Method	ff_return_forward	ff_return_left	ff_return_rot
Ours	**0.088**	**0.087**	**0.104**
AstroLoc	1.065	<u>0.159</u>	0.281
AstroLoc (VIO)	0.269	0.587	0.385
AstroLoc2	<u>0.209</u>	0.245	<u>0.206</u>
AstroLoc2 (VIO)	0.677	0.330	0.477
ORB-SLAM3	0.298	0.145	0.107
DROID-SLAM	0.378	0.142	0.062
VGGT-SLAM	1.416	2.673	1.703
COLMAP†	0.101	0.090	0.096
LiMAP†	0.079	0.062	0.087

† Evaluated on sequences with significant illumination changes.

map localization), and LiMAP [22] (3D line mapping). In contrast, the proposed method performs monocular localization without pre-built maps or appearance features. ATE [23] is computed after Umeyama alignment [24] following [5].

A. Astrobee ISS Free-Flyer Datasets

Table I summarizes the ATE results on the Astrobee sequences. Our method achieves the best accuracy across all trajectories compared to Astrobee-specific localization methods, AstroLoc and AstroLoc2. On average, our method reduces translation error by 66.7% over AstroLoc and 57.3% over AstroLoc2. Fig. 3 shows estimated poses well aligned in the global frame defined by the hatch centers.

Compared to modern SLAM systems, our method reduces ATE by up to 70.5% over ORB-SLAM3 and 76.7% over DROID-SLAM. VGGT-SLAM, a hybrid system that integrates a feed-forward 3D reconstruction backbone with a SLAM backend, also struggles on the ISS sequences. This is likely due to the large domain gap between its training data and ISS imagery, along with the absence of loop closure and global bundle adjustment for long-sequence drift correction. As shown in Fig. 4, its reconstructed map becomes globally inconsistent, leading to large trajectory errors.

B. Illumination Changes on Astrobee Sequences

To assess robustness under varying lighting conditions, we construct an illumination-change evaluation by separating the Astrobee sequences into two phases: mapping images captured under bright lighting and query images under dark lighting, as shown in Fig. 4. This evaluation is conducted only for COLMAP and LiMAP, as they are the most map-dependent methods and thus the most relevant baselines for assessing robustness under illumination changes.

As shown in Table I, our method achieves a mean ATE of 0.093 m, comparable to COLMAP (0.099 m) and LiMAP (0.076 m). Both methods rely on a pre-built feature map that is computationally expensive and must be rebuilt when the environment changes. In contrast, our method exploits conic primitives and Manhattan-world lines as purely geometric cues invariant to illumination. These structural cues enable stable localization under both bright and dark conditions, ensuring consistent performance across all sequences.

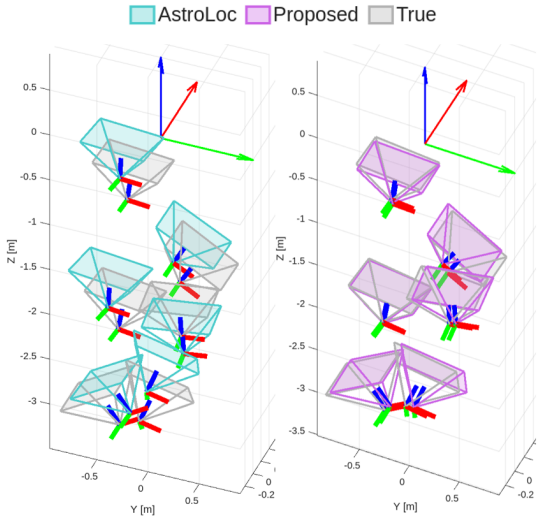


Fig. 3. Localization results of AstroLoc (Cyan) and the proposed method (magenta) in the global hatch frame, compared to ground truth (gray).

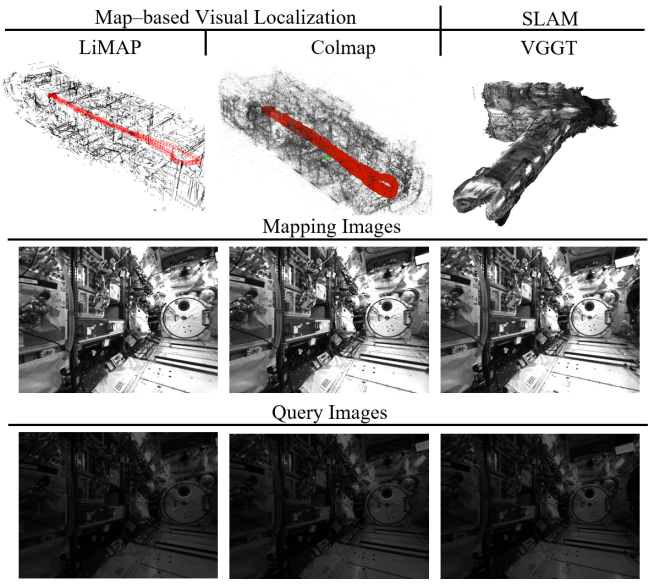


Fig. 4. Comparison on the ISS Astrobee dataset: map-based visual localization under illumination changes with bright mapping images and dark query images, and hybrid feed-forward SLAM on the same sequences.

C. Runtime Analysis

The proposed method is implemented in MATLAB R2024a on an Intel Core i5-12400F (2.50 GHz) desktop. Detection (LSD, YOLO ellipse detector, AAMED fitting) requires $\sim$36 ms per frame, the P2C solver $\sim$8 ms, and CBA $\sim$43 ms, suggesting suitability for near real-time operation.

IV. CONCLUSION

We present a map-free monocular localization method for free-flying robots on the ISS, leveraging purely geometric cues from two circular landmarks without pre-built maps or learned appearance features. Experiments on Astrobee datasets show accuracy comparable to map-based methods with stronger robustness under illumination changes.

V. Limitation

The proposed method has several limitations that we plan to address in future work. First, it requires at least two co-visible circular landmarks with known radii in every frame, limiting it to structured environments such as ISS modules. Second, the P2C solver assumes that the two circles lie on coplanar planes, as is the case for the hatch and window in ISS modules. It therefore does not generalize to circular landmarks placed in arbitrary orientations.

Despite these constraints, the method is applicable to any environment where circular or elliptical structures with known radii are present and co-visible, such as other space stations, industrial facilities, or man-made structures with circular windows, ports, or signage. The requirements are: (1) known physical radii of at least two circular landmarks, (2) their simultaneous visibility in the image, and (3) approximately parallel normal vectors between the two circles.

To overcome these limitations, future work will extend the framework along two directions. We plan to generalize the P2C formulation to handle circles in more diverse spatial configurations, including non-parallel and skewed arrangements, allowing the method to operate in a wider range of scenes. In addition, we will validate the approach on other free-flying robot platforms to demonstrate its generalization across diverse operational environments.

References

[1] M. G. Bualat, T. Smith, E. E. Smith, T. Fong, and D. Wheeler, "Astrobee: A new tool for iss operations," in *2018 SpaceOps Conference*, 2018, p. 2517.

[2] L. Fluckiger and B. Coltin, "Astrobee robot software: Enabling mobile autonomy on the iss," Tech. Rep., 2019.

[3] T. Smith, O. Alexandrov, J. Barlow, J. Benavides, M. Bualat, R. Carlino, B. Coltin, J. Cortez, E. Daley, J. Feller, *et al.*, "Astrobee: Free-flying robots for the international space station," *IEEE Transactions on Field Robotics*, 2026.

[4] D. Hirano, S. Mitani, T. Nishishita, and S. P. Yamaguchi, "Int-ball2: On-orbit demonstration of autonomous intravehicular flight and docking for image capturing and recharging," *IEEE RAM*, 2024.

[5] S. Kang, R. Soussan, D. Lee, B. Coltin, A. M. Vargas, M. Moreira, Trey, *et al.*, "Astrobee iss free-flyer datasets for space intra-vehicular robot navigation research," *IEEE RA-L*, 2024.

[6] P. Kim, B. Coltin, O. Alexandrov, and H. J. Kim, "Robust visual localization in changing lighting conditions," in *IEEE ICRA*, 2017.

[7] R. Soussan, M. Moreira, B. Coltin, and T. Smith, "Astroloc2: Fast sequential depth-enhanced localization for free-flying robots," in *IEEE ICRA*, 2025.

[8] B. Coltin, J. Fusco, Z. Moratto, and R. Nakamura, "Localization from visual landmarks on a free-flying robot," in *IEEE IROS*, 2016.

[9] V. Schröder, R. Regele, J. Sommer, and C. Karrasch, "Gnc system design for the crew interactive mobile companion," in *IAF Human Spaceflight Symp., Int. Astronaut. Congr.*, 2018.

[10] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE TRO*, 2017.

[11] E. Arnold, J. Wynn, S. Vicente, G. Garcia-Hernando, A. Monszpart, D. Turmukhambetov, and E. Brachmann, "Map-free visual relocalization: Metric pose relative to a single image," in *ECCV*, 2022.

[12] V. Leroy, Y. Cabon, and J. Revaud, "Grounding image matching in 3d with mast3r," in *ECCV*, 2024.

[13] S. Wang, V. Leroy, Y. Cabon, B. Chidlovskii, and J. Revaud, "Dust3r: Geometric 3d vision made easy," in *IEEE CVPR*, 2024.

[14] P. Lindenberger, P.-E. Sarlin, and M. Pollefeys, "Lightglue: Local feature matching at light speed," in *IEEE ICCV*, 2023.

[15] C. Meng, Z. Li, X. Bai, and F. Zhou, "Arc adjacency matrix-based fast ellipse detection," *IEEE T-IP*, 2020.

[16] K. Joo, P. Kim, M. Hebert, I. S. Kweon, and H. J. Kim, "Linear rgb-d slam for structured environments," *IEEE T-PAMI*, 2021.

[17] R. Soussan, V. Kumar, B. Coltin, and T. Smith, "Astroloc: An efficient and robust localizer for a free-flying robot," in *IEEE ICRA*, 2022.

[18] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam," *IEEE T-RO*, 2021.

[19] Z. Teed and J. Deng, "Droid-slam: Deep visual slam for monocular, stereo, and rgb-d cameras," *NeurIPS*, 2021.

[20] D. Maggio, H. Lim, and L. Carlone, "Vggt-slam: Dense rgb slam optimized on the sl (4) manifold," *arXiv preprint arXiv:2505.12549*, 2025.

[21] J. L. Schoenberger and M. Pollefeys, "Colmap-structure-from-motion and multi-view stereo," in *IEEE CVPR*, 2016.

[22] S. Liu, Y. Yu, R. Pautrat, M. Pollefeys, and V. Larsson, "3d line mapping revisited," in *IEEE CVPR*, 2023.

[23] J. Sturm, N. Engelhard, F. Endres, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *IEEE IROS*, 2012.

[24] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE TPAMI*, 2002.